ELSEVIER Theoretical Computer Science 204 (1998) 205–231

Theoretical Computer Science

A construction on finite automata that has remained hidden[1]

Jacques Sakarovitch *

Laboratoire Traitement et Communication de l'Information C.N.R.S., URA 820, Ecole Nationale Supérieure des Télécommunications, 46, rue Barrault, F-75634 Paris, France

Abstract

We show how a construction on matrix representations of two tape automata proposed by Schützenberger to prove that rational functions are unambiguous can be given a central rôle in the theory of relations and functions realized by finite automata, in such a way that the other basic results such as the "Cross-Section Theorem", its dual the theorem of rational uniformisation, or the decomposition theorem of rational functions into sequential functions, appear as direct and formal consequences of it. © 1998 Published by Elsevier Science B.V. All rights reserved

Résumé

Nous montrons comment une construction sur la représentation matricielle des automates à deux bandes proposée par Schützenberger pour prouver que toute fonction rationnelle est non ambiguë est en fait au cœur de la théorie des relations et fonctions réalisées par automates finis et permet d'établir naturellement les autres résultats fondamentaux de la théorie comme le "Cross-Section Theorem", son dual, le théorème d'uniformisation rationnelle ou celui de décomposition des fonctions rationnelles en fonctions séquentielles. © 1998 Published by Elsevier Science B.V. All rights reserved

Keywords: Finite automata; Finite transducers; Rational relations; Unambiguous functions; Cross-section theorem

In 1961, Schützenberger [11] made a "remark" on finite transducers. He first defined a *transducer* to be the composition of what we call now a left sequential function by a right sequential function. And he proved that such mappings from one free monoid into another are closed under composition.

Few years later, in a paper "that received less attention that it deserved"[2] Elgot and Mezei [4] proved that rational relations are closed under composition and, moreover,

* E-mail: sakarovitch@enst.fr.

[1] A preliminary version of this paper has been presented to *the Conference on Semigroups and Applications, Saint Petersburg, June 1995* under the title: *The Schützenberger construct and two applications* cf. also the technical report LITP 96/30.

[2] As wrote Eilenberg in his treatise on automata [3].

PII S0304-3975(98)00040-1

that the transducer defined by Schützenberger is indeed *the* model of computation for the rational functions i.e.

Theorem 1 (Elgot and Mezei [4], Decomposition Theorem). *Any rational function is the product of a left sequential function by a right sequential function.* [3]

To tell the truth, the original proof of Decomposition Theorem in [4] is rather hard to follow. It has thus been completely reworked by Eilenberg and Schützenberger who proved in [3] the result directly on the *bimachines* – that was the new name given to the transducers of [11] since by then the word transducer had been used by other authors with an other meaning. It follows as a corollary [4] of Theorem 1:

Theorem 2 (Eilenberg [3], Unambiguity Theorem). *Any rational function may be realized by an unambiguous 2-automaton.* [5]

In [3], Theorem 2 is obtained as a corollary of another more general result (quoted in Section 4 as *Rational Uniformisation Theorem*) which is itself a consequence of the so-called *Rational Cross-Section Theorem* that is established *via* a purely *ad hoc* proof [3, Theorem X.7.1].

Later, in [14], Schützenberger proposed a new proof of Unambiguity Theorem by means of a construction that establishes that any rational function may be given a *matrix representation* of a certain kind – called *semi-monomial* – which yields immediately unambiguity.

Our purpose in this paper is to explain how and why this construction can be given a central position in all this theory we have just sketched, with the other results being derived from it. We first show that the construction, when applied to rational relations instead of rational functions, is a proof of Rational Uniformisation Theorem. The Rational Cross-Section Theorem is then a consequence of it, and Unambiguity Theorem a corollary as before; a more satisfactory genealogy between results is restored. More important, it appears that the construction itself, or the *semi-monomial* representations it yields, is indeed another way of defining bimachines, directly on their matrix representations and without introducing a new concept of automaton. Moreover, the Decomposition Theorem is *directly read on the semi-monomial representation*, provided few technical adjustments are made beforehand.

This construction thus deserves to be carefully presented. We describe it in the frame work of *covering of automata* – which is derived from the notion of covering of graphs that was proposed by Stallings [16] – and which makes (in our opinion) the whole subject much clearer.

---

[3] The statement as it is given here is not entirely correct; the true one is to be found at Section 5.

[4] Which was not explicitly stated in [4].

[5] Notions such as unambiguous 2-automata, sequential functions or matrix representation, will be defined in the body of the paper.

Let us mention that deriving the Rational Uniformisation Theorem from Schützenberger construction is not new in itself: Arnold and Latteux gave such a presentation [1] together with the observation that both Rational Cross-Section and Decomposition Theorems are then corollaries. If it comes to credits, one word is to be added. Up to our knowledge, Rational Uniformisation Theorem first appeared in 1969 in a paper of Kobayashi [9] (who called it "*Single-valuedness Theorem*") and where Unambiguity Theorem also appeared for the first time.[6] The construction used there remains to be compared with the one presented here.

As a conclusion, let us quote from Schützenberger communication to the IFIP Congress in 1965:

> *Like all applications of mathematics, the theory* [*of automata*] *has* [*the following*] *tasks: classifying the problems, extracting the proper concepts and unifying the arguments;...*

This is what we have tried to do here.

1. Automata, as usual

We basically follow the definitions and notations of [3, 10]. We recall some of them though, which differ from another classical way of defining finite automata [8]. We then remind the reader of matrix representation of automata, for it will be one of our basic tools.

The identity of a monoid M is denoted by 1_M, by 1 if no ambiguity is feared. The set of *words* over a finite alphabet A, *i.e.* the set of *finite sequences* of elements of A, or *the free monoid* over A is denoted by A^*. Its identity, or *empty word* is denoted by 1_{A^*}.

1.1. Automata as labelled graphs

A finite *automaton over a finite alphabet* A, $\mathscr{A} = \langle Q, A, E, I, T\rangle$ is a directed graph labelled by elements of A; Q is the finite set of vertices, called *states*, $I \subset Q$ is the set of *initial* states, $T \subset Q$ is the set of *terminal* states and $E \subset Q \times A \times Q$ is the set of labelled *edges*. We shall consider only *finite automata* and thus call them simply *automata* in the sequel. We also note $p \xrightarrow{a} q$ for $(p,a,q) \in E$, or even $p \xrightarrow[\mathscr{A}]{a} q$ if there is a possible ambiguity on the automaton. A *computation* c in $\mathscr{A}$ is a finite sequence of labelled edges that form a path in the graph:

$$c = p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} p_2 \cdots \xrightarrow{a_n} p_n$$

[6] The core of the paper is devoted to classification of formal languages by means of rational transductions and in that area too the paper seems to have been completely overlooked.

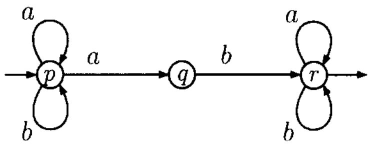

Fig. 1. An automaton $\mathscr{A}_1$ that recognizes the set of words with a factor *ab*.

The *label* of the computation c is the element $a_1 a_2 \cdots a_n$ of A^*. The computation c is *successful* if $p_0 \in I$ and $p_n \in T$. The *behaviour* of $\mathscr{A}$ is the subset $|\mathscr{A}|$ of A^* consisting of labels of successful computations of $\mathscr{A}$.

A state q is said to be *accessible* if there exists a path in $\mathscr{A}$ starting in I and terminating in q. The *accessible part* of $\mathscr{A}$ is the set of its accessible states together with the adjacent edges. A state p is said to be *co-accessible* if there exists a path in $\mathscr{A}$ starting in p and terminating in T. The automaton $\mathscr{A}$ is *trim* if every state is both accessible and co-accessible.

The automaton $\mathscr{A}$ is *complete* if for every state p in Q and every letter a in A there exists *at least one* state q such that (p,a,q) is an edge in E; $\mathscr{A}$ is *deterministic* if for every state p in Q and every letter a in A there exists *at most one* state q such that (p,a,q) is an edge in E. The automaton $\mathscr{A}$ is *unambiguous* if for every *pair* of states (p,q) and every *word* f in A^* there exists *at most one* computation from p to q with label f. A trimmed automaton $\mathscr{A}$ is unambiguous if and only if for every word f in $|\mathscr{A}|$ there exists a unique successful computation with label f.

Example 1. Automata as labelled graphs have natural graphic representation (Fig. 1).

A subset of A^* is said to be *rational* if and only if it is the behaviour of an automaton over A.[7] The family of rational subsets of A^* is denoted by $\mathrm{Rat}\, A^*$.

This definition of automata as labelled graphs extends readily to automata over any monoid: an *automaton over M*, $\mathscr{A} = \langle Q, M, E, I, T\rangle$ is a directed graph the edges of which are labelled by elements of the monoid M. The automaton is finite if the set of edges $E \subset Q \times M \times Q$ is finite (and thus Q is finite). The label of a computation

$$c = p_0 \xrightarrow{x_1} p_1 \xrightarrow{x_2} p_2 \cdots \xrightarrow{x_n} p_n$$

is the element $x_1 x_2 \cdots x_n$ of M. The *behaviour* of $\mathscr{A}$ is the subset $|\mathscr{A}|$ of M consisting of labels of successful computations of $\mathscr{A}$. In this context an automaton *over an alphabet A* is indeed an automaton *over the free monoid A^**.

Two automata, over A or over M, are said to be *equivalent* if they have *the same behaviour*.

[7] This statement is usually considered as a theorem, and not as a definition, for the family of rational subsets is commonly defined as the smallest family of subsets of A^* that contains the finite subsets and that is closed under union, product, and the operation of taking the generated submonoid (cf. [3, 8, 10]). We do not need to refer to this basic result here.

1.2. Matrix representation of automata

Any finite automaton $\mathscr{A}=\langle Q,A,E,I,T\rangle$ may be given a *matrix representation* (λ,μ,ν) over the Boolean semiring $\mathbb{B}$ where $\mu : A^* \to \mathbb{B}^{Q\times Q}$ is the morphism defined by

$$\forall p,q\in Q,\ \forall a\in A \quad a\mu_{p,q}=\begin{cases}1 & \text{if } (p,a,q)\in E,\\ 0 & \text{otherwise}\end{cases}$$

and where λ and ν are, respectively, the row- and column-vectors defined by

$$\forall p\in Q \quad \lambda_p=1 \Leftrightarrow p\in I; \qquad \forall p\in Q \quad \nu_p=1 \Leftrightarrow p\in T.$$

The triple (λ,μ,ν) is a representation of $\mathscr{A}$ in the sense that it allows to compute the elements of the behaviour of $\mathscr{A}$:

$$|\mathscr{A}|=\{f\in A^* \mid (\lambda\cdot f\mu\cdot\nu)=1\}.$$

It is also known that if the entries, 0 and 1, of λ,μ and ν are considered to belong to $\mathbb{N}$ instead to $\mathbb{B}$ – i.e. if (λ,μ,ν) is a $\mathbb{N}$-representation – then, for every f in A^*, $(\lambda\cdot f\mu\cdot\nu)$ is *the number of distinct successful paths* in $\mathscr{A}$ with label f. Hence the automaton $\mathscr{A}$ is *unambiguous* if and only if, for every f in A^*, every entry of $f\mu$ is 0 or 1 (whereas all computations are made in $\mathbb{N}$).

Example 1 (*continued*). The matrix representation of $\mathscr{A}_1$ is

$$\lambda_1=(1\ \ 0\ \ 0),\quad a\mu_1=\begin{pmatrix}1&1&0\\0&0&0\\0&0&1\end{pmatrix},\quad b\mu_1=\begin{pmatrix}1&0&0\\0&0&1\\0&0&1\end{pmatrix},\quad \nu_1=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

If it is considered as a $\mathbb{N}$-representation it follows that

$$(abab)\mu_1=\begin{pmatrix}1&0&2\\0&0&0\\0&0&1\end{pmatrix},$$

where it is read that there are indeed 2 successful paths in $\mathscr{A}_1$ with label *abab*.

Direct product of automata translates into the *tensor product* of their representations. Let us first recall that the tensor product of two matrices X and Y of dimension $P\times Q$ and $R\times S$, respectively, and with entries in a semiring $\mathbb{K}$, is the matrix $X\otimes Y$ of dimension $(P\times R)\times(Q\times S)$ defined by

$$\forall p\in P,\ \forall q\in Q,\ \forall r\in R,\ \forall s\in S, \quad X\otimes Y_{(p,r),(q,s)}=X_{p,q}Y_{r,s}.$$

It is noteworthy that $X\otimes Y$ has a natural *block decomposition* (which will be currently use in the sequel): $X\otimes Y$ is a block-matrix of dimension $P\times Q$ of blocks of dimension $R\times S$ (or *vice versa*). The tensor product of representations makes sense because of the following.

Lemma 1 (Schützenberger [13]). *Let $\mathbb{K}$ be any commutative semiring and M any monoid. Let $\mu : M \to \mathbb{K}^{Q\times Q}$ and $\kappa : M \to \mathbb{K}^{R\times R}$ be two morphisms. The mapping $\mu \otimes \kappa$ defined for every m in M by*

$$m\mu \otimes \kappa = m\mu \otimes m\kappa$$

is a morphism.

We define the tensor product of two representations (λ, μ, ν) and (η, κ, ζ) to be

$$(\lambda, \mu, \nu) \otimes (\eta, \kappa, \zeta) = (\lambda \otimes \eta, \mu \otimes \kappa, \nu \otimes \zeta).$$

It easily follows then from Lemma 1:

Proposition 2 (Schützenberger [13]). *The representation of the direct product of two automata over an alphabet A is the tensor product of the representations of the automata.*

An example of the tensor product of two representations appears in Section 3.2.

2. Covering

The notion of *covering* as defined by Stallings [16] for graphs can be extended to automata (since automata are, as we said, labelled graphs) and proved to be perfectly suited to deal with the constructions on automata we are aiming at. Its presentation has been already partly published in [6]; it is made more complete here.

2.1. Morphism of automata

Given an automaton $\mathscr{A} = \langle Q, M, E, I, T\rangle$, the set E of labelled edges is canonically equipped with three mappings (the three projections):

$$\iota : E \to Q, \quad \tau : E \to Q, \quad \text{and} \quad \varepsilon : E \to M.$$

The vertices $e\iota$ and $e\tau$ are respectively the *origin* and the *end* of the edge e; $e\varepsilon$ is the *label* of the edge e.

A morphism φ from *an automaton* $\mathscr{B} = \langle R, M, F, J, U\rangle$ *into an automaton* $\mathscr{A} = \langle Q, M, E, I, T\rangle$ is indeed a pair of mappings (both denoted by φ): one between the set of states $\varphi : R \to Q$, and one between the set of edges $\varphi : F \to E$, which satisfy the three properties: [8]

$$\varphi \circ \iota = \iota \circ \varphi \quad \text{and} \quad \varphi \circ \tau = \tau \circ \varphi, \tag{1}$$

$$\varphi \circ \varepsilon = \varepsilon, \tag{2}$$

$$J\varphi \subseteq I \quad \text{and} \quad U\varphi \subseteq T. \tag{3}$$

[8] Though we use the postfixed notation for functions (e.g. $e\iota$) we find it clearer to indicate composition of functions explicitly by a symbol ($\circ$) than with the mere concatenation.

Conditions (1) imply that *the image of a path in* $\mathscr{B}$ *is a path in* $\mathscr{A}$. Condition (2) implies that *the label of a path in* $\mathscr{B}$ *is the same as the label of the image of that path in* $\mathscr{A}$. Conditions (3) imply that *the image of a successful path in* $\mathscr{B}$ *is a successful path in* $\mathscr{A}$. In particular, if $\varphi : \mathscr{B} \to \mathscr{A}$ it holds $|\mathscr{B}| \subseteq |\mathscr{A}|$.

Example 2. The classical construction of *direct product* of automata (over a free monoid $\mathscr{A}$) gives a common and useful instance of morphism of automata. The direct product of $\mathscr{A} = \langle Q, A, E, I, T\rangle$ and $\mathscr{B} = \langle R, A, F, J, U\rangle$ is by definition the automaton $\mathscr{A} \times \mathscr{B} = \langle Q \times R, A, G, I \times J, T \times U\rangle$ where the set G of labelled edges is defined by

$$G = \{((p,r), a, (q,s)) \mid (p,a,q) \in E,\ (r,a,s) \in F\}.$$

The projections $\pi_{\mathscr{A}}$ and $\pi_{\mathscr{B}}$ from the set $Q \times R$ on the first and on the second components respectively, together with the corresponding mappings from G into E and F – i.e. $(p,r)\pi_{\mathscr{A}} = p$, $((p,r),a,(q,s))\pi_{\mathscr{A}} = (p,a,q)$, and so on – are clearly morphisms from $\mathscr{A} \times \mathscr{B}$ on $\mathscr{A}$ and $\mathscr{B}$, respectively.

Example 3. The canonical mapping of a deterministic automaton onto its minimal automaton is a morphism.

2.2. *Covering of automata*

For every state q of an automaton $\mathscr{A} = \langle Q, M, E, I, T\rangle$, let us denote by $\mathrm{Out}_{\mathscr{A}}(q)$ the set [9] of edges of $\mathscr{A}$ *the origin of which is* q, that is edges that are "going out" of q:

$$\mathrm{Out}_{\mathscr{A}}(q) = \{e \in E \mid e\iota = q\}.$$

One defines dually $\mathrm{In}_{\mathscr{A}}(q)$ as the set of edges of $\mathscr{A}$ *the end of which is* q, that is edges that are "going in" q:

$$\mathrm{In}_{\mathscr{A}}(q) = \{e \in E \mid e\tau = q\}.$$

If φ is a morphism from $\mathscr{B} = \langle R, M, F, J, U\rangle$ into $\mathscr{A} = \langle Q, M, E, I, T\rangle$ then for every r in R, φ maps $\mathrm{Out}_{\mathscr{B}}(r)$ into $\mathrm{Out}_{\mathscr{A}}(r\varphi)$, and $\mathrm{In}_{\mathscr{B}}(r)$ into $\mathrm{In}_{\mathscr{A}}(r\varphi)$.

We say that φ is *Out-surjective* (resp. *Out-bijective*, *Out-injective*) if for every r in R the restriction of φ to $\mathrm{Out}_{\mathscr{B}}(r)$ is surjective onto $\mathrm{Out}_{\mathscr{A}}(r\varphi)$ (resp. bijective between $\mathrm{Out}_{\mathscr{B}}(r)$ and $\mathrm{Out}_{\mathscr{A}}(r\varphi)$, injective). Accordingly, we say that φ is *In-surjective* (resp. *In-bijective*, *In-injective*) if for every r in R the restriction of φ to $\mathrm{In}_{\mathscr{B}}(r)$ is surjective onto $\mathrm{In}_{\mathscr{A}}(r\varphi)$ (resp. bijective between $\mathrm{In}_{\mathscr{B}}(r)$ and $\mathrm{In}_{\mathscr{A}}(r\varphi)$, injective).

What we call *Out-bijective morphism* is exactly what Stallings calls a *covering* (*of graphs*). The definition of *covering of automata* we are now coining is consistent with the one of covering of graphs and puts also in relation the initial states and the terminal states, respectively.

[9] Stallings denotes it "$\mathrm{Star}_{\mathscr{A}}(q)$". As the star is the common denomination for the generated submonoid, we cannot keep it, though it nicely conveys the idea of "a set of edges going out" of q.

Definition 1. A morphism φ from an automaton $\mathscr{B} = \langle R, M, F, J, U \rangle$ into an automaton $\mathscr{A} = \langle Q, M, E, I, T \rangle$ is a covering if the following conditions hold:
(i) φ is *Out-bijective*;
(ii) for every i in I, there exists a *unique* j in J such that $j\varphi = i$;
(iii) for every t in T, $t\varphi^{-1} \subset U$ (i.e. by (3) $T\varphi^{-1} = U$).

Example 3 (*continued*). The morphism of a deterministic automaton onto its minimal automaton is a covering.

We also need the dual definition:

Definition 2. A morphism φ from $\mathscr{B}$ into $\mathscr{A}$ is a co-covering if the following conditions hold:
(i) φ is *In-bijective*;
(ii) for every i in I, $i\varphi^{-1} \subset J$ (i.e. by (3) $I\varphi^{-1} = J$);
(iii) for every t in T, there exists a unique s in S such that $s\varphi = t$.

These definitions are set up in view of the following.

Proposition 3. *Any covering (resp. any co-covering) $\varphi : \mathscr{B} \to \mathscr{A}$ induces a bijection between the successful pathes in $\mathscr{B}$ and those in $\mathscr{A}$.*

Proof. Since φ is a morphism, the image $c = d\varphi$ of a successful path d in $\mathscr{B}$ is a successful path in $\mathscr{A}$. It remains thus to show that for every successful path c in $\mathscr{A}$ there exists a *unique* successful path d in $\mathscr{B}$ such that $d\varphi = c$.

The proof is by induction on the length of c (we give it for the case where φ is a *covering*; the dual case is analoguous). We show indeed a slightly more general property:

> (P1) *for every path c in $\mathscr{A}$* the origin of which is an initial state *there exists a unique path d in $\mathscr{B}$* the origin of which is an initial state *and such that $d\varphi = c$.*

Property (P1) holds for $|c| = 0$ since for every i in I there exists a *unique* j in J such that $j\varphi = i$ (Definition 1(ii)). Let $c : i \xrightarrow[\mathscr{A}]{f} p \xrightarrow[\mathscr{A}]{m} q$ a path in $\mathscr{A}$ where (p, m, q) is an edge in $\mathscr{A}$. By induction hypothesis, there exists a *unique path* $e : j \xrightarrow[\mathscr{B}]{f} s$ such that $e\varphi = i \xrightarrow[\mathscr{A}]{f} p$ (thus $s\varphi = p$) and such that j is an initial state (in $\mathscr{B}$). Since φ is a covering, there exists a *unique edge* $s \xrightarrow[\mathscr{B}]{m} t$ with origin s and the image of which by φ is $p \xrightarrow[\mathscr{A}]{m} q$ (Definition 1(i)). The path $d : j \xrightarrow[\mathscr{B}]{f} s \xrightarrow[\mathscr{B}]{m} t$ is uniquely determined and satisfies (P1).

If moreover c is a successful path, i.e. if q is a final state, then t, which belongs to $q\varphi^{-1}$, is also a final state (Definition 1(iii)) and d is a successful path. $\square$

Corollary 4. *If $\varphi : \mathscr{B} \to \mathscr{A}$ is a covering (resp. a co-covering) then $\mathscr{B}$ is equivalent to $\mathscr{A}$.*

Corollary 5. *A trimmed covering (resp. co-covering) of a trim unambiguous automaton is an unambiguous automaton.* [10]

Proof. Since it is trim, $\mathscr{A}$ is unambiguous if and only if for every f in $|\mathscr{A}|$ there exists one successful computation with label f. Thus a trim covering of an unambiguous automaton is unambiguous since there is a bijection between the successful computations in the two automata. □

A particular case we get: *A co-covering of a deterministic automaton is unambiguous.*

The last definition we need is the one of *immersion*.

Definition 3. A morphism φ from $\mathscr{B}$ into $\mathscr{A}$ is an *immersion* if the following conditions hold:
(i) φ is *Out-injective*;
(ii) for every i in I there exists *at most one* j in J such that $j\varphi = i$.

Roughly speaking an immersion is a covering from which some edges have been removed and where some states have lost the property of being initial or terminal.

If $\varphi : \mathscr{B} \to \mathscr{A}$ is an immersion it is not only true that $|\mathscr{B}| \subseteq |\mathscr{A}|$ – which holds as soon as there exists a morphism from $\mathscr{B}$ into $\mathscr{A}$ – but φ is moreover *an injection* from the set of successful pathes of $\mathscr{B}$ into the set of successful pathes of $\mathscr{A}$.

Example 4. A *subautomaton* $\mathscr{B}$ of $\mathscr{A}$, that is an automaton obtained from $\mathscr{A}$ by deleting edges and/or by *suppressing the quality of being initial or terminal* to certain states is an immersion (the morphism being the identity mapping on the set of states).

It will be convenient to say that $\mathscr{B}$ *covers* $\mathscr{A}$ or is a *covering of* $\mathscr{A}$ (resp. is an *immersion in* $\mathscr{A}$) if there exists a morphism $\varphi : \mathscr{B} \to \mathscr{A}$ that is a covering (resp. an immersion).

3. The Schützenberger construct

In the case of automata over a free monoid, that is automata that can be determinised by the subset method, a canonical construction allows – as stated in Theorem 4 below – to associate to any automaton $\mathscr{A}$ a particular covering that we call the Schützenberger covering, or S-covering of $\mathscr{A}$. That covering is the first step of a construction that yields the following result.

[10] The statement would hold indeed without the assumption the automaton being trim. But the proof would be less direct then.

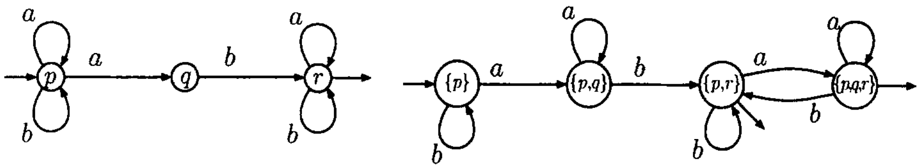

Fig. 2. The automaton $\mathscr{A}_1$ and its determinised by the subset method $\mathscr{A}_{1\mathscr{D}}$.

Theorem 3. *Let $\mathscr{A}$ be an automaton on A^*. Then there exists an unambiguous automaton that is equivalent to $\mathscr{A}$ and that is an immersion in $\mathscr{A}$.*

The essence of this statement lies of course in the fact that the quoted unambiguous automaton is at the same time *equivalent to* and an *immersion in* $\mathscr{A}$. For otherwise, the deterministic automaton $\mathscr{A}_{\mathscr{D}}$ associated to $\mathscr{A}$ by the subset construction is obviously unambiguous and equivalent to $\mathscr{A}$; but it cannot be immersed in $\mathscr{A}$: there is no relationships between the pathes in $\mathscr{A}$ and those in $\mathscr{A}_{\mathscr{D}}$, as it can be observed for instance on Fig. 2 [The edge $(\{p,q\},a\{p,q\})$ of $\mathscr{A}_{1\mathscr{D}}$ cannot be given an image in any mapping from $\mathscr{A}_{1\mathscr{D}}$ onto $\mathscr{A}_1$ in order to make a morphism.]

The immersion we get is a subautomaton of the S-covering of $\mathscr{A}$. We present the construction in two different frameworks: on the automata as labelled graphs, and on the matrix representations.

3.1. The construct on labelled graphs

As we just did, we note $\mathscr{A}_{\mathscr{D}}$ the deterministic automaton obtained from an automaton $\mathscr{A}$ over a free monoid by the subset method (and we call it the *determinised* of $\mathscr{A}$).

Theorem & Definition 4. *Let $\mathscr{A}$ be an automaton and $\mathscr{A}_{\mathscr{D}}$ its determinised. Let $\mathscr{S}$ be the accessible part of $\mathscr{A}_{\mathscr{D}} \times \mathscr{A}$. It then holds:*
(i) *$\pi_{\mathscr{A}}$ is a covering of $\mathscr{S}$ onto $\mathscr{A}$.*
(ii) *$\pi_{\mathscr{A}_{\mathscr{D}}}$ is an In-surjective morphism from $\mathscr{S}$ onto $\mathscr{A}_{\mathscr{D}}$.*
We call $\mathscr{S}$ the S-covering *of $\mathscr{A}$.*

Example 1 (*continued*). The S-covering of $\mathscr{A}_1$ is shown on Fig. 3.

In order to prove Theorem 4, we first establish two properties of morphisms of automata on a free monoid. In the sequel, $\mathscr{A} = \langle Q,A,E,I,T\rangle$ and $\mathscr{B} = \langle R,A,F,J,U\rangle$ are two automata on A^*.

Property 1. Let $\mathscr{B}$ be a *deterministic* and *complete* automaton on A. For any automaton $\mathscr{A}$ on A, $\pi_{\mathscr{A}}$ is an *Out-bijective* morphism from $\mathscr{B} \times \mathscr{A}$ onto $\mathscr{A}$.

Proof. Let us keep the notations of Example 2. For every (r,p) in $R \times Q$ we have

$$\mathrm{Out}_{\mathscr{B}\times\mathscr{A}}(r,p) = \{((r,p),a,(s,q)) \mid (r,a,s) \in F,\ (p,a,q) \in E\}.$$

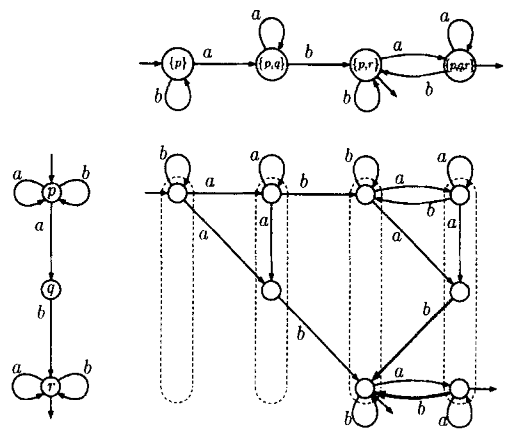

Fig. 3. The S-covering of $\mathcal{A}_1$.

By hypothesis on $\mathcal{B}$, for every a in A there exists, r being fixed, *exactly one edge* (r,a,s) in F. There exists then exactly one edge $((r,p),a,(s,q))$ in G for every edge (p,a,q) in $E:\pi_{\mathcal{A}}$ is a bijection between $\mathrm{Out}_{\mathcal{B}\times\mathcal{A}}(p,r)$ and $\mathrm{Out}_{\mathcal{A}}(p)$. $\square$

Property 2. Let $\varphi:\mathcal{B}\to\mathcal{A}$ be an *Out-bijective* morphism and let $\mathcal{C}$ be the *accessible part* of $\mathcal{B}$. Then the restriction of φ to $\mathcal{C}$ is an *Out-bijective* morphism from $\mathcal{C}$ onto $\mathcal{A}$.

Proof. Since $\mathcal{C}$ is obtained from $\mathcal{B}$ by deleting some states, and the edges that arrive in and start from those deleted states, it might be the case that $\mathrm{Out}_{\mathcal{C}}(r)$ were strictly contained in $\mathrm{Out}_{\mathcal{B}}(r)$ and thus that the restriction of φ to $\mathcal{C}$ induces only an *injection* from $\mathrm{Out}_{\mathcal{C}}(r)$ into $\mathrm{Out}_{\mathcal{A}}(r\varphi)$ and not a bijection. But if r in R is accessible in $\mathcal{B}$ the states $\{s \mid (r,m,s)\in F\}$ are all accessible and

$$\mathrm{Out}_{\mathcal{C}}(r)=\{f \mid f_\iota = r \text{ and } f\tau \text{ accessible}\}=\mathrm{Out}_{\mathcal{B}}(r). \qquad \square$$

A more general statement yields condition (i) of Theorem 4 as a particular case.

Proposition 6. *Let $\mathcal{A}$ be an automaton, $\mathcal{B}$ a deterministic automaton equivalent to $\mathcal{A}$, and $\mathcal{S}$ the accessible part of $\mathcal{B}\times\mathcal{A}$. Then $\pi_{\mathcal{A}}$ is a covering from $\mathcal{S}$ onto $\mathcal{A}$.*

Proof. By Properties 1 and 2, $\pi_{\mathcal{A}}$ is an *Out-bijective* morphism from $\mathcal{S}$ onto $\mathcal{A}$.

Since $\mathcal{B}$ (as any deterministic automaton) has only one initial state $J=\{r_0\}$, then for every initial state i of $\mathcal{A}$ there exists one and only one initial state in $i\pi_{\mathcal{A}}^{-1}:(r_0,i)$.

Let now (r,p) in $R\times Q$ an accessible state of $\mathcal{B}\times\mathcal{A}$, *i.e.* there exists f in A^* and i in I such that

$$(r_0,i)\xrightarrow[\mathcal{B}\times\mathcal{A}]{f}(r,p) \quad \text{thus} \quad r_0\xrightarrow[\mathcal{B}]{f}r \quad \text{and} \quad i\xrightarrow[\mathcal{A}]{f}p.$$

If p is in T, then f is in $|\mathcal{A}|$ and r is in U since $\mathcal{B}$ is equivalent to $\mathcal{A}$: all the states of $\mathcal{S}$ that are mapped onto p by $\pi_{\mathcal{A}}$ are terminal.

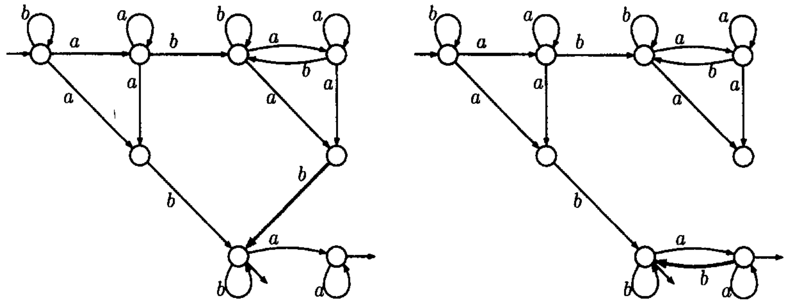

Fig. 4. The two S-immersions in $\mathscr{A}_1$.

The three conditions for being a covering have been checked for $\pi_{\mathscr{A}}: \mathscr{S} \to \mathscr{A}$. $\square$

Proof of Theorem 4. It remains to prove condition (ii).

Let $\mathscr{A}_{\mathscr{D}} = \langle 2^Q, A, F, J, U \rangle$.[11] By definition,

$$F = \{(P,a,S) \in 2^Q \times A \times 2^Q \mid S = \{s \mid \exists p \in P \ (p,a,s) \in E\}\},$$

$$J = \{I\} \quad \text{and} \quad U = \{S \in 2^Q \mid S \cap T \neq \emptyset\}.$$

From this definition follows:

$$P \xrightarrow[\mathscr{A}_{\mathscr{D}}]{a} S \Leftrightarrow S = \{q \mid \exists p \in P \ p \xrightarrow[\mathscr{A}]{a} q\}$$

and then

$$\begin{aligned} \forall P, S \subset Q, \ \forall q \in S \quad P \xrightarrow[\mathscr{A}_{\mathscr{D}}]{a} S &\Rightarrow \exists p \in P \ p \xrightarrow[\mathscr{A}]{a} q \\ &\Rightarrow \exists p \in P \ (P,p) \xrightarrow[\mathscr{A}_{\mathscr{D}} \times \mathscr{A}]{a} (S,q) \end{aligned}$$

which expresses that $\pi_{\mathscr{A}_{\mathscr{D}}}: \mathscr{S} \to \mathscr{A}_{\mathscr{D}}$ is In-surjective. $\square$

Proof of Theorem 3. Let $\mathscr{S}$ be the S-covering of an automaton $\mathscr{A}$. Since $\pi_{\mathscr{A}_{\mathscr{D}}}$ est In-surjective from $\mathscr{S}$ onto $\mathscr{A}_{\mathscr{D}}$, it is possible, by deleting some edges in $\mathscr{S}$ if $\pi_{\mathscr{A}_{\mathscr{D}}}$ is not In-injective, and by suppressing if necessary their quality of being terminal to certain states, to construct a sub-automaton $\mathscr{T}$ of $\mathscr{S}$ that is a *co-covering* of $\mathscr{A}_{\mathscr{D}}$. Such a $\mathscr{T}$ is thus unambiguous, and equivalent to $\mathscr{A}_{\mathscr{D}}$ and thus to $\mathscr{A}$. Since $\mathscr{S}$ is a covering of $\mathscr{A}$, $\mathscr{T}$ is an immersion in $\mathscr{A}$. $\square$

Example 1 (*continued*). In the case of the S-covering of $\mathscr{A}_1$, there is only one state, namely $(\{p,r\},r)$, where $\pi_{\mathscr{A}_{1\mathscr{D}}}$ is not In-bijective; there are thus two possible sub-automata, as shown on Fig. 4, that are immersions equivalent to $\mathscr{A}_1$.

[11] This should not be confusing, for $\mathscr{A}_{\mathscr{D}}$ and $\mathscr{B}$ never appear in the same statement; on the contrary, $\mathscr{A}_{\mathscr{D}}$ happens to be a special case of an automaton $\mathscr{B}$.

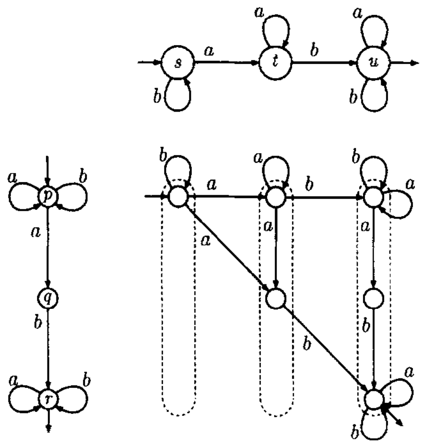

Fig. 5. Another covering of $\mathscr{A}_1$.

The virtue of the Schützenberger construct will probably be even better understood by considering a "non-example". Fig. 5 shows the direct product of $\mathscr{A}_1$ with its minimal deterministic equivalent automaton $\mathscr{B}_1$. In the state (u, q), $\pi_{\mathscr{B}_1}$ is not In-surjective.

3.2. The construct on matrix representations

The above construct may be rephrased in terms of matrix representations. In itself, this does not bring in anything new. But the framework of representations proves to be better suited for one of the applications we have in mind: the decomposition theorem for rational functions.

Let (λ, μ, ν) be the representation of $\mathscr{A} = \langle Q, A, E, I, T \rangle$ and (η, κ, ξ) the representation of $\mathscr{A}_{\mathscr{D}} = \langle 2^Q, A, F, J, U \rangle$, i.e.

$$\forall P, S \in Q, \ \forall a \in A \quad a\kappa_{P,S} = 1 \Leftrightarrow S = \{q \mid \exists p \in P \ a\mu_{p,q} = 1\},$$

$$\eta_S = 1 \Leftrightarrow S = \{q \mid \lambda_q = 1\} \quad \text{and} \quad \xi_P = 1 \Leftrightarrow \exists p \in P \ \nu_p = 1.$$

By definition, $a\kappa$ is *row-monomial*, i.e. every line has at most one non-zero entry (this is clearly equivalent to the fact that $\mathscr{A}_{\mathscr{D}}$ is deterministic).

By Proposition 2 the representation of $\mathscr{A}_{\mathscr{D}} \times \mathscr{A}$ is $(\eta, \kappa, \xi) \otimes (\lambda, \mu, \nu)$. Any matrix $(f)\kappa \otimes \mu$ is a $2^Q \times 2^Q$ block matrix made of blocks of size $Q \times Q$.

In order to describe the representation of the S-covering, the accessible part of $\mathscr{A}_{\mathscr{D}} \times \mathscr{A}$, we need another notation. Let α be any $Q \times R$-matrix (over any semiring $\mathbb{K}$) and let P be any subset of Q. We denote by $\alpha^{[P]}$ the matrix the lines of which are equal to those of α if their index is in P and to 0 otherwise, i.e.

$$\forall q \in R \quad (\alpha^{[P]})_{p,q} = \begin{cases} \alpha_{p,q} & \text{if } p \in P, \\ 0 & \text{otherwise.} \end{cases}$$

The dimension of the representation (ζ,σ,ω) of the S-covering is the one of $(\eta,\kappa,\xi)\otimes(\lambda,\mu,\nu)$: $2^Q\times Q$. For every a in A, the matrix $a\sigma$ is a $2^Q\times 2^Q$ block matrix obtained by replacing *the* non-zero entry of the line P of $a\kappa$ by the $Q\times Q$-matrix $a\mu^{[P]}$, i.e.

$$a\sigma_{(P,Q),(S,Q)}=\begin{cases} a\mu^{[P]} & \text{if } a\kappa_{P,S}=1,\\ 0 & \text{otherwise.}\end{cases}$$

Accordingly,

$$\zeta_{(S,Q)}=\begin{cases} \lambda & \text{if } \eta_S=1,\\ 0 & \text{otherwise,}\end{cases}$$

and

$$\forall P\subset Q \quad \omega_{(P,Q)}=\nu^{[P]}.$$

Example 1 (*continued*). The matrix representation of $\mathscr{A}_1$ is

$$\lambda_1=(1\quad 0\quad 0),\qquad a\mu_1=\begin{pmatrix}1&1&0\\0&0&0\\0&0&1\end{pmatrix},\qquad b\mu_1=\begin{pmatrix}1&0&0\\0&0&1\\0&0&1\end{pmatrix},\qquad \nu_1=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

The matrix representation of $\mathscr{A}_{1\mathscr{D}}$ is

$$\eta_1=(1\quad 0\quad 0\quad 0),\qquad a\kappa_1=\begin{pmatrix}0&1&0&0\\0&1&0&0\\0&0&0&1\\0&0&0&1\end{pmatrix},$$

$$b\kappa_1=\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\end{pmatrix},\qquad \xi_1=\begin{pmatrix}0\\0\\1\\1\end{pmatrix}.$$

The representation $(\zeta_1,\sigma_1,\omega_1)$ of the S-covering $\mathscr{S}_1$ of $\mathscr{A}_1$ is then

$$\zeta_1=(1\quad 0\quad 0\quad\ 0\quad 0\quad 0\quad\ 0\quad 0\quad 0\quad\ 0\quad 0\quad 0),$$

$$a\sigma_1=\begin{pmatrix}0&\begin{smallmatrix}1&1&0\\0&0&0\\0&0&0\end{smallmatrix}&0&0\\ 0&\begin{smallmatrix}1&1&0\\0&0&0\\0&0&0\end{smallmatrix}&0&0\\ 0&0&0&\begin{smallmatrix}1&1&0\\0&0&0\\0&0&1\end{smallmatrix}\\ 0&0&0&\begin{smallmatrix}1&1&0\\0&0&0\\0&0&1\end{smallmatrix}\end{pmatrix},\qquad b\sigma_1=\begin{pmatrix}\begin{smallmatrix}1&0&0\\0&0&0\\0&0&0\end{smallmatrix}&0&0&0\\ 0&0&\begin{smallmatrix}1&0&0\\0&0&1\\0&0&0\end{smallmatrix}&0\\ 0&0&\begin{smallmatrix}1&0&0\\0&0&0\\0&0&1\end{smallmatrix}&0\\ 0&0&\begin{smallmatrix}1&0&0\\0&0&1\\0&0&1\end{smallmatrix}&0\end{pmatrix},$$

$$\omega_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

where the "big zeroes" represent 3×3-blocks of zeroes.

The reader will easily check that this is the representation of the automaton shown on Fig. 3, with the difference that, in order to have all blocks of the same size (3×3), every dashed box in the figure is supposed to contain 3 states, the missing ones being the initial or terminal state of no edge whatsoever. □

The definition of the representation of a S-immersion from (ζ, σ, ω) is then straightforward. For every letter a in A, every non-zero block $a\sigma_{(P,Q),(S,Q)}$ is replaced by a *column-monomial* block which has the same non-zero columns as the original one. In other words, every non-zero non-monomial column of any $Q \times Q$-block of $a\sigma$ is made column-monomial, *but not zero*, by the deletion of arbitrary entries. The same operation is performed on the $Q \times 1$-block vectors of ω.

The easiest way to prove that such a representation yields an automaton equivalent to $\mathscr{A}$ is probably to come back to labelled graphs (an In-surjective morphism that is made In-bijective).

Example 1 (*continued*). There are two S-immersions $\mathscr{T}_1'$ and $\mathscr{T}_1''$ in $\mathscr{A}_1$ with representations $(\zeta_1, \sigma_1', \omega_1')$ and $(\zeta_1, \sigma_1'', \omega_1'')$, respectively. Obviously $a\sigma_1' = a\sigma_1'' = a\sigma_1$ and $\omega_1' = \omega_1'' = \omega_1$. The construction has a real impact on $b\sigma_1'$ and $b\sigma_1''$:

$$b\sigma_1' = \begin{pmatrix} \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & \boxed{0} \\ 0 & 0 & 1 \end{matrix} & 0 \end{pmatrix}, \quad b\sigma_1'' = \begin{pmatrix} \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \boxed{0} \end{matrix} & 0 \end{pmatrix},$$

where the "deleted" entries have been marked by a box for easier finding.

The example shows clearly that it is *the blocks* of $a\sigma$ (or $b\sigma$) that are transformed into column-monomial blocks, not the whole matrix $a\sigma$ (or $b\sigma$). We shall come back to this point later, at Section 4.

4. Uniformisation of rational relations

We now consider automata over direct products of free monoids or *2-automata*. The behaviour of such an automaton $\mathscr{C} = \langle Q, A^* \times B^*, E, I, T\rangle$ is a subset of $A^* \times B^*$, that is the graph of a relation θ from A^* into B^*:

$$\forall f \in A^* \quad f\theta = \{g \in B^* \mid (f,g) \in |\mathscr{C}|\}.$$

A relation from A^* into B^* is said to be *rational* if and only if it is the behaviour of an automaton over $A^* \times B^*$.[12] A rational relation, or the automaton $\mathscr{B}$ that realizes it, is *unambiguous* if any element of $|\mathscr{B}|$ is the label of a *unique* successful computation in $\mathscr{B}$.

Definition 4 (Uniformisation of a relation). Let $\theta : A^* \to B^*$ be any relation; a function $\tau : A^* \to B^*$ is said to uniformise θ, (or to be a uniformisation of θ), if it selects one element in $f\theta$ for every $f \in \operatorname{Dom}\theta$.

In other words, τ is a function such that $\operatorname{Dom}\tau = \operatorname{Dom}\theta$ and for any f in $\operatorname{Dom}\theta$, $f\tau$ is in $f\theta$.

Our aim is to establish the following result as a direct consequence of the Schützenberger construct.

Theorem 5 (Kobayashi [9, Theorem 2], Eilenberg [3, Proposition IX.8.2]; Rational Uniformisation Theorem). *Any rational relation is uniformised by an unambiguous rational function.*

We start from the characterization of rational relations by their matrix representation.

Theorem 6 (Eilenberg [3, Theorem IX.5.1]; Berstel [2, Theorem III.7.1]; Kleene–Schützenberger Theorem). *A relation $\theta : A^* \to B^*$ is rational if and only if it exists a representation (λ, μ, ν) over the semiring* $\operatorname{Rat} B^*$ *such that for every f in A^**

$$f\theta = \lambda \cdot f\mu \cdot \nu.$$

Going back from representation to automaton, Theorem 6 implies – or, indeed, is equivalent to say – that a rational relation $\theta : A^* \to B^*$, represented by (λ, μ, ν), may be

[12] What was said in footnote 7 for subsets of A^* obviously applies to subsets of $A^* \times B^*$.

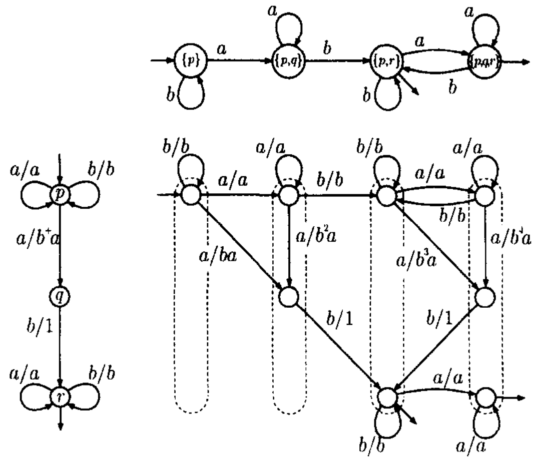

Fig. 6. The 2-automaton $\mathscr{C}_1$ and a S-uniformisation of θ_1.

realized by a finite automaton $\mathscr{C}$ every edge of which is labelled by a pair $(a, a\mu_{p,q})$ with a in A and $a\mu_{p,q}$ in Rat B^*. Such an automaton $\mathscr{C}$ is sometimes called a *real-time* transducer for it "reads" one letter at every transition.

Definition 5. The *underlying input automaton* of a 2-automaton $\mathscr{C} = \langle Q, A^* \times B^*, F, I, T \rangle$ is the (1-)automaton $\mathscr{A} = \langle Q, A^*, E, I, T \rangle$ obtained from $\mathscr{C}$ by erasing the second component of the label of every edge.

An automaton $\mathscr{C}$ is *real-time* if and only if its underlying input automaton is a "classical" automaton over A i.e. the label of every transition is a letter in A.

4.1. *Proof (of Theorem 5) on labelled graphs*

Let (λ, μ, ν) be a representation of θ and let $\mathscr{C}$ and $\mathscr{A}$ as above. Let $\mathscr{T}$ be the automaton given by Theorem 3 (starting from $\mathscr{A}$). Every edge (r, a, s) in $\mathscr{T}$ corresponds – since $\mathscr{T}$ is an immersion in $\mathscr{A}$ – to a unique edge (p, a, q) in $\mathscr{A}$ and thus to a unique edge $(p, (a, a\mu_{p,q}), q)$ in $\mathscr{C}$. If we choose, arbitrarily, one word w in $a\mu_{p,q}$ (w thus depends on $a, r,$ and s) we can build from $\mathscr{T}$ a 2-automaton $\mathscr{U}$ by replacing every edge (r, a, s) by $(r, (a, w), s)$. The relation τ realized by $\mathscr{U}$ is an unambiguous function – for $\mathscr{T}$ is unambiguous – has the same domain as θ – for $\mathscr{T}$ is equivalent to $\mathscr{A}$ – and its graph is contained in the one of θ – by the choice of w. $\square$

Example 1 (*continued*). Let θ_1 be the relation from $\{a, b\}^*$ into $\{a, b\}^*$ that replaces in any word one of its factor ab by a word in b^+a. Fig. 6 shows an automaton $\mathscr{C}_1$ that realises θ_1, the underlying input automaton of which is $\mathscr{A}_1$ (on the left, vertically) and the result of the construction described in the above proof.

4.2. *Matrix representation of* S-*uniformisation*

Let (λ, μ, ν) be a representation – of dimension Q – of θ and, as in the previous section, let $\mathscr{C}$ be the 2-automaton defined by (λ, μ, ν) and $\mathscr{A}$ its underlying input automaton. Let (η, κ, ξ) be the (Boolean) representation – of dimension 2^Q – of $\mathscr{A}_{\mathscr{D}}$. As in Section 3.2, the representation (ζ, σ, ω) of the S-covering of $\mathscr{C}$ is defined by

$$\forall P, S \subset Q \quad a\sigma_{(P,Q),(S,Q)} = \begin{cases} a\mu^{[P]} & \text{if } a\kappa_{P,S} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Accordingly,

$$\forall S \subset Q \quad \zeta_{(S,Q)} = \begin{cases} \lambda & \text{if } \eta_S = 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\forall P \subset Q \quad \omega_{(P,Q)} = \nu^{[P]}.$$

The making of the representation of a S-uniformisation goes then as in Section 3.2: for every letter a in A, every non-zero block $a\sigma_{(P,Q),(S,Q)}$ is replaced by a *column-monomial* block which has the same non-zero columns as the original one. In other words, every non-zero non-monomial column of any $Q \times Q$-block of $a\sigma$ is made column-monomial, *but not zero*, by replacing every non-empty entry (thus a subset of B^*) either by zero or by an arbitrary single element that belongs to it. The same operation is performed on the $Q \times 1$-block vectors of ω.

Example 1 (*continued*). The matrix representation of $\mathscr{C}_1$ is

$$\lambda_1 = (1 \quad 0 \quad 0), \qquad a\mu_1 = \begin{pmatrix} a & b^+a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a \end{pmatrix},$$

$$b\mu_1 = \begin{pmatrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & b \end{pmatrix}, \qquad \nu_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The representation $(\zeta_1, \sigma_1, \omega_1)$ of the S-covering of $\mathscr{C}_1$ is then

$$\zeta_1 = (1 \;\; 0 \;\; 0 \quad 0 \;\; 0 \;\; 0 \quad 0 \;\; 0 \;\; 0 \quad 0 \;\; 0 \;\; 0),$$

$$a\sigma_1 = \begin{pmatrix} 0 & \begin{matrix} a & b^+a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} a & b^+a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 \\ 0 & 0 & 0 & \begin{matrix} a & b^+a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a \end{matrix} \\ 0 & 0 & 0 & \begin{matrix} a & b^+a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a \end{matrix} \end{pmatrix}, \quad b\sigma_1 = \begin{pmatrix} \begin{matrix} b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & b \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & b \end{matrix} & 0 \end{pmatrix},$$

$$\omega_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The representation $(\zeta_1, \sigma_1', \omega_1')$ of the S-uniformisation of θ_1 shown in Fig. 6 is given by

$$a\sigma_1' = \begin{pmatrix} 0 & \begin{matrix} a & ba & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} a & b^2a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 \\ 0 & 0 & 0 & \begin{matrix} a & b^3a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a \end{matrix} \\ 0 & 0 & 0 & \begin{matrix} a & b^4a & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a \end{matrix} \end{pmatrix}, \quad b\sigma_1' = \begin{pmatrix} \begin{matrix} b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & b \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 \end{pmatrix},$$

and $\omega_1' = \omega_1$.

The representation of the S-uniformisation we have built has thus a characteristic property that will play such a rôle in the sequel that it deserves to be properly stated.

Definition 6 (*Schützenberger* [14]). A matrix m is said to be (*left–right*) *semi-monomial* if there exists a block decomposition of m such that m is row-monomial as a matrix of blocks and that every non-zero block in m is column-monomial.

Accordingly, a representation (λ, μ, ν) is said to be *(left–right) semi-monomial* if all matrices λ, ν, and $a\mu$ for every letter a, are semi-monomial matrices with congruent block decomposition.

The Schützenberger construct on matrix representation can then be summed up as

Proposition 7. *Any rational relation can be uniformised by a function with a semi-monomial representation.*

Furthermore, it can be specialized for functions as

Proposition 8 (Schützenberger [14]). *Any rational function has a semi-monomial representation.*

5. Decomposition of rational functions

A rational function is said to be *(left) sequential* if the underlying input automaton of a 2-automaton that realizes it is *deterministic* with *all states being terminal*; dually, a rational function is said to be *right sequential* if the underlying input automaton of a 2-automaton that realizes it is *co-deterministic* with *all states being initial.* The left sequential functions are those realized by the automata called *generalized sequential machines* (*gsm*) in [7, 3] and *are probably among the oldest concept in formal language theory.* [13]

In order to give the correct statement to Theorem 1, let us call *proper* a function (or a relation) from a free monoid into a free monoid with the property that the image of the empty word is the empty word.

Theorem 1 (Elgot and Mezei [4]). *Any proper rational function may be obtained as the composition of a left sequential function by a right sequential function.*

The condition of being proper is necessary for any sequential function (left or right) is proper and so is the composition of two proper functions.

The Schützenberger construct as seen at Section 3.2 yields a matrix representation of functions on which the decomposition may be seen directly, provided one knows how the sequential functions are represented and how the composition of functions translates onto the representations, which we recall first.

5.1. Representation of (sub-)sequential functions

It follows from the definition that a rational function $\alpha : A^* \to B^*$ is left sequential if and only if it has a representation (λ, μ, ν) with the property that

[13] Quoted from [2].

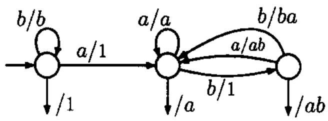

Fig. 7. The automaton $\mathscr{C}_1$.

(i) for every a in A, $a\mu$ is a *row-monomial* matrix, and every non-zero entry consists in a single word of B^*;

(ii) λ has only one non-zero entry, which is equal to 1;

(iii) every entry of ν is 1.

Dually, a rational function $\alpha : A^* \to B^*$ is right sequential if and only if it has a representation (λ, μ, ν) with the property that

(i) for every a in A, $a\mu$ is a *column-monomial* matrix, and every non-zero entry consists in a single word of B^*;

(ii) every entry of λ is 1;

(iii) ν has only one non-zero entry, which is equal to 1.

As Schützenberger did later (cf. [15]), it will be useful to consider slightly more general families of functions, namely the left and right *subsequential* functions, easily defined by their representations.

A rational function $\alpha : A^* \to B^*$ is *(left) subsequential* if it has a *row-monomial* representation, i.e. as above, $a\mu$ is a *row-monomial* matrix, with every non-zero entry consisting in a single word of B^*, λ has only one non-zero entry, which is equal to 1, but there are no other condition on ν than every non-zero entry consists in a single word of B^*. Thus, a subsequential function differs from a sequential one by the fact that some states of an automaton that realizes it may not be terminal and that to those which are terminal is attached a word of B^*, namely the corresponding non-zero entry of ν, that is suffixed to the output of any path that ends at that state.

Example 5. Let $\mathscr{C}_1$ be the automaton over $A^* \times A^* = \{a,b\}^* \times \{a,b\}^*$ that replaces the factors *abb* of any word by factors *baa* which realizes that substitution[14] while reading the word from left to right, the last a of a substituted factor being combined with a possible following factor *bb* in order to give rise to a new substitution. This automaton is represented in Fig. 7.

The matrix representation of $\mathscr{C}_1$ is

$$\lambda_1 = (1 \quad 0 \quad 0), \qquad a\mu_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & a & 0 \\ 0 & ab & 0 \end{pmatrix},$$

[14] It is called "Fibonacci substitution" for it corresponds to the numerical equivalence in the numeration system defined by the sequence of Fibonacci numbers, when a is interpreted as the digit 0 and b as 1 (cf. [5] for more details on the subject).

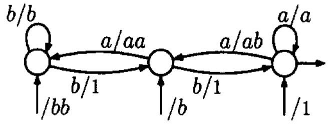

Fig. 8. The automaton $\mathscr{D}_1$.

$$b\mu_1 = \begin{pmatrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & ba \end{pmatrix}, \qquad \nu_1 = \begin{pmatrix} 1 \\ a \\ ab \end{pmatrix}.$$

Dually, a rational function $\alpha : A^* \to B^*$ is *right subsequential* if it has a *column-monomial* representation, i.e. as above, $a\mu$ is a *column-monomial* matrix, with every non-zero entry consisting in a single word of B^*, ν has only one non-zero entry, which is equal to 1, but there are no other conditions on λ than that every non-zero entry consists in a single word of B^*.

Example 5 (*continued*). Let us now consider the function from A^* into itself that is obtained by replacing the factors *abb* of any word by factors *baa* and but that realizes the substitution *while reading the word from right to left*, the last b of a substituted factor being combined with a possible following factor *ab* in order to give rise to a new substitution. The automaton $\mathscr{D}_1$ that realizes this function by reading the word *from left to right* (as usual) is represented in Fig. 8.

The matrix representation of $\mathscr{D}_1$ is

$$\eta_1 = (bb \quad b \quad 1), \qquad a\kappa_1 = \begin{pmatrix} 0 & 0 & 0 \\ aa & 0 & 0 \\ 0 & ab & a \end{pmatrix},$$

$$b\kappa_1 = \begin{pmatrix} b & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \xi_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

5.2. Representation of the composition product

It is also due to Elgot and Mezei [4] that *the (composition) product of two rational relations is a rational relation.* Here again, the original proof is not easy and a classical proof goes by a so-called "diamond lemma" (cf. [2, 3] for instance) after the rational relations have been characterized as the product of inverse morphism, intersection with a rational set and direct morphism.

But the closure under product may also be based on the representation of rational relations by matrices with rational entries [11].

Let $\alpha : A^* \to B^*$ and $\beta : B^* \to C^*$ be two rational relations with representations (λ, μ, ν) and (η, κ, ξ) respectively. Consider first the two morphisms μ and κ; and let us give a meaning to the expression $(f\mu)\kappa$ for any f in A^*: we shall take the image of every entry of $f\mu$ by κ and we thus form a block matrix. The following statement shows that this construction is legitimate.

Lemma 9 (Schützenberger [11]). *Let $\mu : A^* \to (\mathrm{Rat}\, B^*)^{Q\times Q}$ and $\kappa : B^* \to (\mathrm{Rat}\, C^*)^{R\times R}$ be two morphisms (of monoids). Then the mapping*

$$\pi : A^* \to (\mathrm{Rat}\, C^*)^{(Q\times R)\times(Q\times R)}$$

defined by

$$f\pi_{(q,R),(s,R)} = (f\mu_{q,s})\kappa \tag{4}$$

is a morphism.

The morphism π thus defined from μ and κ is called the *Kronecker product* of μ by κ and denoted by

$$\pi = \mu \circ \kappa.$$

(Beware: this product, in contrast with tensor product, is a *non-commutative* one.)

The Kronecker product of (λ, μ, ν) by (η, κ, ξ) is the representation

$$(\chi, \pi, \psi) = (\lambda, \mu, \nu) \circ (\eta, \kappa, \xi)$$

defined by

$$\pi = \mu \circ \kappa, \ \forall q \in Q \quad \chi_{(q,R)} = \eta \cdot (\lambda_q)\kappa \quad \text{and} \quad \forall q \in Q \quad \psi_{(q,R)} = (\nu_q)\kappa \cdot \xi. \tag{5}$$

We can then state

Proposition 10 (Schützenberger [11]). *Let $\alpha : A^* \to B^*$ and $\beta : B^* \to C^*$ be two rational relations. The Kronecker product of a representation of α by a representation of β is a representation of the (composition) product $\alpha\beta$.*

Example 5 (*continued*). The Kronecker product of $(\lambda_1, \mu_1, \nu_1)$ by $(\eta_1, \kappa_1, \xi_1)$, $(\chi_1, \pi_1, \psi_1) = (\lambda_1, \mu_1, \nu_1) \circ (\eta_1, \kappa_1, \xi_1)$ is thus

$$\chi_1 = (bb \quad b \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0),$$

$$a\pi_1 = \begin{pmatrix} 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & 0 & 0 \\ aa & 0 & 0 \\ 0 & ab & a \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & 0 & 0 \\ aab & aa & 0 \\ 0 & 0 & ab \end{matrix} & 0 \end{pmatrix}, \qquad b\pi_1 = \begin{pmatrix} \begin{matrix} b & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} & 0 & 0 \\ 0 & 0 & \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \\ 0 & 0 & \begin{matrix} aa & 0 & 0 \\ 0 & ab & a \\ 0 & 0 & 0 \end{matrix} \end{pmatrix},$$

$$\psi_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ a \\ 0 \\ 0 \\ ab \end{pmatrix}.$$

Since the Kronecker product of two row-monomial matrices (resp. two column-monomial matrices) is clearly a row-monomial matrix (resp. a column-monomial matrix), Proposition 10 immediately yields:

Corollary 11. *The composition product of two left* (*resp. right*) *subsequential functions is a left* (*resp. right*) *subsequential function.*

The same observation together with the direct application of (5) gives

Corollary 12. *The composition product of two left* (*resp. right*) *sequential functions is a left* (*resp. right*) *sequential function.*

And finally:

Corollary 13. *The composition product of a left subsequential function by a right subsequential function has a semi-monomial representation.*

5.3. Decomposition of semi-monomial representations

Based on Proposition 10, the first and main step toward Theorem 1 is given by the following converse of Corollary 13.

Proposition 14. *Let* $\tau : A^* \to B^*$ *be a rational function with a semi-monomial representation* (χ, π, ψ). *Then there exist*

(i) *an alphabet* Z,

(ii) *a left subsequential function* $\theta : A^* \to Z^*$ *with a row-monomial representation* (λ, μ, ν), *and*

(iii) *a right subsequential function* $\sigma : Z^* \to B^*$ *with a column-monomial representation* (η, κ, ξ), *such that* $(\chi, \pi, \psi) = (\lambda, \mu, \nu) \circ (\eta, \kappa, \xi)$ *holds* (*and thus* $\tau = \theta\sigma$).

Proof. Let (χ, π, ψ) be a semi-monomial representation *of dimension* $Q \times R$, i.e., for every a in A, $a\pi$ is a $Q \times Q$ block matrix, every block being an $R \times R$ matrix.

Let then $Z = \{Q \times A \times Q\} \cup \{Q\}$. The representation (λ, μ, ν) is defined by the following:

$$\mu : A^* \to (\operatorname{Rat} Z^*)^{Q \times Q},$$

$$\forall p,q \in Q,\ \forall a \in A \quad a\mu_{p,q} = \begin{cases} (p,a,q) & \text{if the block } a\pi_{(p,R),(q,R)} \text{ is non-zero,} \\ 0 & \text{otherwise} \end{cases}$$

Since $a\pi$ is semi-monomial, $a\mu$ is row-monomial, for every a in A. Furthermore,

$$\forall q \in Q \quad \lambda_p = \begin{cases} 1 & \text{if the block } \chi_{(q,R)} \text{ is non-zero,} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\forall p \in Q \quad v_p = p.$$

And the representation (η, κ, ξ) is defined by the following:

$$\kappa : Z^* \to (\operatorname{Rat} B^*)^{R \times R},$$

$$\forall p,q \in Q,\ \forall a \in A \quad (p,a,q)\kappa = a\pi_{(p,R),(q,R)}.$$

Since $a\pi$ is semi-monomial, $(p,a,q)\kappa$ is column-monomial, for every (p,a,q) in Z. Let now r be a fixed element in R, arbitrarily chosen. For every p in Q, $p\kappa$ is an $R \times R$ matrix, the rth column of which is equal to $\psi_{(p,R)}$ and with all other columns equal to 0. Since ψ is semi-monomial, $p\kappa$ is column-monomial, for every p in Z. The two vectors η and ξ are then defined by

$$\eta = \chi_{(q,R)} \quad \text{where } \chi_{(q,R)} \text{ is the non zero block of } \chi,$$

$$\forall s \in R \quad \xi_s = \begin{cases} 1 & \text{if } s = r, \\ 0 & \text{otherwise.} \end{cases}$$

It follows then directly from (4) and (5) that $(\chi, \pi, \psi) = (\lambda, \mu, v) \circ (\eta, \kappa, \xi)$ and thus, by Proposition 10, $\tau = \theta\sigma$. □

Two technical steps will lead us from the proof of Proposition 14 to the one of Theorem 1.

Proof of Theorem 1. We keep the notations of the preceding proof. Let p_- be the only state in Q such that λ_{p_-} is not zero (and thus equal to one). The first step consists in changing θ into a sequential function: let $\theta' : A^* \to Z^*$ be the function with representation (λ, μ, v') where v' is such that $v'_p = 1$ for every p in Q. By definition, θ' is sequential. If $f = a_1 a_2 \cdots a_n$ is a word of A^*, there exists a unique sequence $p_1, p_2, \ldots, p_n$ of states in Q such that

$$f\theta = (p_-, a_1, p_1)(p_1, a_2, p_2) \cdots (p_{n-1}, a_n, p_n) p_n;$$

and we then have

$$f\theta' = (p_-, a_1, p_1)(p_1, a_2, p_2) \cdots (p_{n-1}, a_n, p_n).$$

With a construction adapted from a classical one (cf. [12]), we define a subsequential mapping $\sigma' : Z^* \to B^*$ such that $\theta'\sigma' = \theta\sigma = \tau$. Let t be a new state not in R and $R' = R \cup \{t\}$. The representation (η', κ', ξ') is defined by the following:

$$\kappa' : Z^* \to (\mathrm{Rat}\, B^*)^{R' \times R'}$$

$$\forall p, q \in Q,\ \forall a \in A \quad (p,a,q)\kappa' = \begin{pmatrix} \boxed{(p,a,q)\kappa} & \boxed{\cdot} \\ \boxed{\quad 0 \quad} & 0 \end{pmatrix} \quad \longleftarrow v_{(p,a,q)}$$

where

$$v_{(p,a,q)} = (p,a,q)\kappa \cdot \psi_q$$

is a monomial (column-)vector and

$$\eta' = \begin{pmatrix} \boxed{\quad \eta \quad} & 0 \end{pmatrix} \quad \text{and} \quad \xi' = \begin{pmatrix} \boxed{0} \\ 1 \end{pmatrix}.$$

It is straightforward to check that for every f in A^*, $\eta \cdot f\kappa \cdot \xi = \eta' \cdot f\kappa' \cdot \xi'$.

The second step is cleared by two remarks that are easy exercises. Let \$ be a symbol not in Z and $Z' = Z \cup \{\$\}$. On one hand, the function $\theta'' : A^* \to Z'^*$ defined by $f\theta'' = \$f\theta'$ for every f in A^* (i.e. $f\theta'' = \$(p_-, a_1, p_1)(p_1, a_2, p_2) \cdots (p_{n-1}, a_n, p_n)$ if $f = a_1 a_2 \cdots a_n$) is clearly a left sequential function. On the other hand, the function $\sigma'' : Z'^* \to B^*$ defined by $(\$u)\sigma'' = u\sigma'$ is a right *sequential* function since σ' is subsequential (cf. [2, Exercise IV.2.1]). Thus $\theta''\sigma'' = \theta'\sigma' = \theta\sigma = \tau$. □

References

[1] A. Arnold, M. Latteux, A new proof of two theorems about rational transductions, Theoret. Comput. Sci. 8 (1979) 261–263.

[2] J. Berstel, Transductions and Context-Free Languages, Teubner, Stuttgart, 1979.

[3] S. Eilenberg, Automata, Languages and Machines, vol. A, Academic Press, New York, 1974.

[4] C.C. Elgot, J.E. Mezei, On relations defined by generalized finite automata, IBM J. Res. Dev. 9 (1965) 47–68.

[5] Ch. Frougny, Fibonacci representations and finite automata, IEEE Trans. Inform. Theory 37 (1991) 393–399.

[6] Ch. Frougny, J. Sakarovitch, Synchronized rational relations of finite and infinite words, Theoret. Comput. Sci. 108 (1993) 45–82.

[7] S. Ginsburg, The Mathematical Theory of Context-Free Languages, Addison-Wesley, Reading, MA, 1966.

[8] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, Reading, MA, 1979.
[9] K. Kobayashi, Classification of formal languages by functional binary transductions, Inform. and Control 15 (1969) 95–109.
[10] D. Perrin, Finite Automata, in: J. van Leeuwen (Ed.), Handbook of Theoretical Computer Science, Elsevier, Amsterdam, 1990, pp. 1–53.
[11] M.P. Schützenberger, A remark on finite transducers, Inform. and Control 4 (1961) 381–388.
[12] M.P. Schützenberger, Certain elementary families of automata, in: Proc. Symp. on Mathematical Theory of Automata, Polytechnic Institute of Brooklyn, 1962, pp. 139–153.
[13] M.P. Schützenberger, On a theorem of R. Jungen, Proc. Amer. Math. Soc. 13 (1962) 885–890.
[14] M.P. Schützenberger, Sur les relations rationnelles entre monoïdes libres, Theoret. Comput. Sci. 3 (1976) 243–259.
[15] M.P. Schützenberger, Une variante des fonctions séquentielles, Theoret. Comput. Sci. 4 (1977) 47–57.
[16] J.R. Stallings, Topology of finite graphs, Invent. Math. 71 (1983) 551–565.